

High prevalence of *MAP2K1* mutations in variant and IGHV4-34 expressing hairy-cell leukemia

Joshua J. Waterfall¹, Evgeny Arons², Robert L. Walker¹, Marbin Pineda¹, Laura Roth²,
J. Keith Killian¹, Ogan D. Abaan¹, Sean R. Davis¹, Robert J. Kreitman², Paul S. Meltzer^{1*}

¹Genetics Branch and ²Laboratory of Molecular Biology, Center for Cancer Research,
NCI, NIH, Bethesda MD 20892 USA.

Supplementary Note

Mutation and Sample Purity Analysis

Reads were aligned using bwa¹ version 0.5.9 to the hg19 human reference genome and duplicate reads were marked with picard².

Initial non-reference variant calls were performed for each sample (tumor and normal separately) with the samtools package². For each tumor/normal pair the set of non-reference positions were joined and the exome data were re-queried for variant allele frequency (alternate read fraction) at each site. Germline heterozygous SNPs that underwent loss of heterozygosity in the tumor provide quantitative estimates of both tumor (pre-treatment blood) and normal (post-treatment blood) samples purity. Theoretically they would have variant allele frequency 0.5 in the normal sample and variant allele frequency 0 or 1 in the tumor (e.g. from loss of the alternate or reference allele respectively) and the degree of mixture between these two states can be solved for from the observed variant allele frequencies. For instance, for SNPs where the reference allele was lost in the tumor the formulas are:

$$v_{obs}^{normal} = \lambda_{normal}(0.5) + (1 - \lambda_{normal})(1)$$

$$v_{obs}^{tumor} = \lambda_{tumor}(1) + (1 - \lambda_{tumor})(0.5)$$

where v_{obs}^{normal} is the observed variant allele frequency in the normal, λ_{normal} is the fraction of normal DNA in the normal sample (i.e. the normal sample purity) and similarly for the tumor. Individuals RB31, BL26, BL14, 10984, RG06, and RG01 have both sufficiently pure tumor and normal ($\lambda \geq 0.8$ in each sample) to allow somatic variant identification at the current depth of coverage (Supplementary Figure 1). Individuals BL42, HH14, 10748, and 10821 had high tumor contamination in the normal and therefore the normal samples were not included in the following analyses.

For the six individuals with sufficiently pure tumor and matched normal samples (see above) somatic variants were identified with the samtools package² using the paired sample mode to identify mutations specifically in the tumor. In the remaining four individuals the matched normal sample was too contaminated with tumor (see above) and therefore only the tumor was analyzed for non-reference variants. Variants across all samples were then pooled and annotated with annovar³ and in house scripts. These annotations were used to filter against 1000 Genomes⁴, the NHLBI ESP5400⁵, our database of in house normal samples sequenced in the Meltzer lab and segmental duplications⁶ to filter out probable germline variants, while retaining any variants identified in COSMIC⁷. To identify driver genes we further filtered these remaining variants for those predicted to affect the encoded protein (i.e. nonsynonymous substitutions, both frameshift and non-frameshift indels, and splice site mutations) and which occur at conserved nucleotides as identified by phastCons⁸. All such variants are listed in Supplementary Table 4. Finally, the count of such variants for each gene was used to identify genes mutated in more than one individual, insisting that at least one such

mutation be clearly somatic (i.e. in one of the six samples where this could be discerned).

Supplementary References

1. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–60 (2009).
2. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–9 (2009).
3. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
4. Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
5. Tennesen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–9 (2012).
6. Bailey, J. A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–7 (2002).
7. Forbes, S. A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945–50 (2011).
8. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–50 (2005).

Supplementary Table 1. Patient and tumor characteristics for all samples included in this study. D+ denotes dim staining and B+ denotes bright staining.

Patient ID	Age	Sex	Diagnosis	CD25	CD11c	CD103	CD123	Anxa1	VH	Homology
10731	73	M	HCLv	-	+	+			VH4-61	91.43
10748	54	M	HCL	B+	B+	+			VH4-34	98.74
10812	55	M	HCLv	-	B+	+			VH4-34	99.58
10821	76	M	HCLv	-	+	+	D+		VH4-39	100
10872	75	F	HCLv	-	B+	+	-		VH1-46	93.8
10888	70	M	HCLv	D+	+	-	-		VH4-34	100
10920	46	M	HCL	+	B+	+	+		VH3-30	97.11
10984	69	M	HCLv	-	B+	+	-		VH3-15	96.77
90061	49	M	HCLv	-	B+	+			VH3-48	99.59
BH05	46	M	HCL	+	B+	+	+		VH3-30	95.87
BH14	58	M	HCL	B+	B+	+			VH3-9	91.8
BH18	57	M	HCL	B+	B+	+		WK-	VH4-34	99.58
BH21	40	M	HCL	+	B+	+	+		VH5-51	95.45
BH28	42	M	HCLv	-	B+	+			VH4-34	99.58
BH29	55	M	HCLv	-	B+	+	+		VH4-34	100
BH31	69	M	HCL	B+	B+	+			VH3-33	97.93
BH34	56	M	HCL	+	B+	D+			VH4-34	98.74
BH36	58	M	HCL	B+	B+	+	+		VH3-23	96.28
BL02	39	F	HCL	+	+	+			VH1-2	95.87
BL14	43	M	HCLv	D+	+	+	D+		VH4-31	95.51
BL18	69	M	HCLv	-	B+	+			VH4-34	100
BL26	55	M	HCLv	-	-	-			VH4-34	99.18
BL29	37	M	HCL	B+	B+	+			VH3-33	94.63
BL42	80	M	HCL	+	B+	B+		NEG	VH4-34	99.58
BL43	44	M	HCL	+	B+	D+			VH3-11	96.69
BL44	39	F	HCL	+	B+	+			VH3-30	94.21
BL46	40	M	HCL	B+	B+	+			VH4-39	95.51
BL47	36	M	HCL	+	B+	+			VH3-7	94.63
HC07	68	M	HCL	B+	B+	+	+		VH4-34	99.16
HH10	45	F	HCLv	-	+	+	-		VH1-18	87.6
HH14	50	M	HCL	+	B+	+			VH4-34	99.58
HH19	44	M	HCLv	-	B+	+	D+		VH5-a	96.28
HH38	54	M	HCL	B+	B+	+	+		VH3-21	97.22
L235	52	M	HCL	B+	B+	+			VH1-2	94.21
RB18	51	M	HCL	B+	+	+	-		VH1-18	97.11
RB24	60	F	HCL	B+	B+	+	-		VH4-34	99.58
RB26	46	M	HCL	B+	B+	+	+		VH4-39	96.41
RB27	50	M	HCL	B+	B+	+	+		VH1-69	92.15

RB31	73	M	HCLv	-	B+	+	-		VH1-8	96.31
RB32	71	M	HCL	B+	B+	+	+		VH1-8	95.87
RC08	45	F	HCL	B+	B+	+	+		VH3-30	95.87
RD15	30	M	HCL	B+	B+	+	+	POS	VH2-70	96.69
RG01	37	M	HCLv	-	B+	+	-	NEG	VH6-1	87.65
RG03	49	F	HCLv	-	B+	+	-		VH3-9	95.45
RG05	53	F	HCLv	-	B+	+	-	NEG	VH4-34	100
RG06	58	F	HCLv	-	+	+	-	NEG	VH4-34	100
RG07	80	M	HCLv	-	+	+	-		VH3-30	100
RG08	69	M	HCLv	-	+	+	+	NEG	VH1-69	99.17
RG09	68	M	HCLv	-	B+	+	-		VH4-4	99.17
RG10	73	F	HCLv	-	B+	+	-	NEG	VH1-46	97.13
RG11	74	M	HCLv	-	B+	+	-	NEG	VH4-34	99

Supplementary Table 2. Quality metrics for ten tumor/normal pairs of whole exome sequencing results for discovery sample set as calculated with picard package². All samples were enriched on the Agilent 50MB Whole Exome bait set.

SAMPLE	BAIT TERRITORY (BASES)	TOTAL READS	PASS FILTER UNIQUE BASES ALIGNED	MEAN BAIT COVERAGE	ZERO COVERAGE TARGETS FRACTION	FRACTION TARGET BASES AT LEAST 10X
RB31 normal	51756122	125911662	9938743982	73.304528	0.030768	0.925701
RB31 tumor	51756122	126088074	9785446710	65.958927	0.031959	0.913501
BL14 normal	51756122	64816920	5221121523	29.034362	0.039702	0.8329
BL14 tumor	51756122	89032212	7237777272	39.275197	0.036396	0.869481
BL26 normal	51756122	60777020	4974554246	29.386309	0.039871	0.831965
BL26 tumor	51756122	61596682	5123819919	29.186218	0.039509	0.837513
BL42 normal	51756122	67621126	5023419333	27.223242	0.040508	0.814592
BL42 tumor	51756122	54972482	4599694981	29.212687	0.040569	0.844168
HH14 normal	51756122	73319778	5813723408	33.825029	0.038905	0.852587
HH14 tumor	51756122	51045390	3990904444	24.862676	0.043636	0.794339
10821 normal	51756122	142744732	10159996263	64.765365	0.031017	0.914421
10821 tumor	51756122	130618722	9634170318	53.169899	0.033788	0.891724
10748 normal	51756122	133240932	9828417014	67.050026	0.030566	0.915359
10748 tumor	51756122	114223500	8305391066	52.115394	0.032222	0.900221
RG01 normal	51756122	73474842	5896644376	33.738677	0.03739	0.856063
RG01 tumor	51756122	67766628	5454009066	30.37965	0.040283	0.83627
RG06 normal	51756122	78075912	5653294907	26.334585	0.038529	0.809484
RG06 tumor	51756122	96408590	7023212737	38.724056	0.033896	0.881348
10984 normal	51756122	185209250	12931710992	84.233586	0.028883	0.936835
10984 tumor	51756122	205017836	14544090703	89.859652	0.029375	0.93421

Supplementary Table 3. *MAP2K1*, *U2AF1*, and *BRAF* mutations in 31 variant and IGHV4-34 expressing hairy-cell leukemia patients and 20 IGHV4-34 negative classic hairy-cell leukemia patients. Whole exome sequencing indicates initial discovery set of 10 samples and Sanger/Taqman sequencing indicates Sanger sequencing for *MAP2K1* and *U2AF1* and Taqman assay for *BRAF* V600E in separate validation set samples.

Patient ID	variant/classic	IGHV4-34	sequencing	MEK1	BRAF	U2AF1
BH14	classic	-	Sanger/Taqman	F53L		
10920	classic	-	Sanger/Taqman		V600E	
BH05	classic	-	Sanger/Taqman		V600E	
BH21	classic	-	Sanger/Taqman		V600E	
BH31	classic	-	Sanger/Taqman		V600E	
BL02	classic	-	Sanger/Taqman		V600E	
BL29	classic	-	Sanger/Taqman		V600E	
BL43	classic	-	Sanger/Taqman		V600E	
BL44	classic	-	Sanger/Taqman		V600E	
BL46	classic	-	Sanger/Taqman		V600E	
BL47	classic	-	Sanger/Taqman		V600E	
HH38	classic	-	Sanger/Taqman		V600E	
L235	classic	-	Sanger/Taqman		V600E	
RB18	classic	-	Sanger/Taqman		V600E	
RB26	classic	-	Sanger/Taqman		V600E	
RB32	classic	-	Sanger/Taqman		V600E	
RC08	classic	-	Sanger/Taqman		V600E	
RD15	classic	-	Sanger/Taqman		V600E	
RB27	classic	-	Sanger/Taqman		NA	
BH36	classic	-	Sanger/Taqman			
BH18	classic	+	Sanger/Taqman	F53L		
BL42	classic	+	exome	I103N		
BH34	classic	+	Sanger/Taqman	K57N		
HH14	classic	+	exome	K57N		
HC07	classic	+	Sanger/Taqman	Q56P		
10748	classic	+	exome		V600E	
RB24	classic	+	Sanger/Taqman		V600E	
BL14	variant	-	exome	42_57del		
90061	variant	-	Sanger/Taqman	C121S		
RB31	variant	-	exome	C121S		
RG03	variant	-	Sanger/Taqman	C121S		
10984	variant	-	exome	F53L		S34F
10872	variant	-	Sanger/Taqman	K57T		
HH19	variant	-	Sanger/Taqman		V600E	
RG10	variant	-	Sanger/Taqman			S34F

10731	variant	-	Sanger/Taqman			
RB06	variant	-	Sanger/Taqman			
10821	variant	-	exome			
RG01	variant	-	exome			
RG07	variant	-	Sanger/Taqman			
RG08	variant	-	Sanger/Taqman			
RG09	variant	-	Sanger/Taqman			
BL18	variant	+	Sanger/Taqman	C121S		
RG05	variant	+	Sanger/Taqman	C121S		
BH29	variant	+	Sanger/Taqman	K57E		
BH28	variant	+	Sanger/Taqman	K57N		
10888	variant	+	Sanger/Taqman		NA	
BL26	variant	+	exome			S34F
10812	variant	+	Sanger/Taqman			
RG06	variant	+	exome			
RG11	variant	+	Sanger/Taqman			

Supplementary Table 4. Candidate driver mutations from whole exome sequencing. Somatic variants from samples RB31, BL26, BL14, 10984, RG06, and RG01 were combined with all non-reference variants from samples BL42, HH14, 10748, and 10821. See Supplementary Note for filtering details. (attached)

Supplementary Table 5. Primers for Sanger sequencing of *MAP2K1* exons 2 and 3 and *U2AF1* exon 2. Each sequencing primer was tailed with M13 sequences (F: tgtaaaacgacggccagt, R: caggaaacagctatgacc). Labels for Taqman assay were VIC (reporter 1) and FAM (reporter 2).

Target	Primer 1	Primer 2
<i>MAP2K1</i> exon 2	TGACTTGTGCTCCCCACTTT	TCCCCAGGCTTCTAAGTACC
<i>MAP2K1</i> exon 2	GACTTGTGCTCCCCACTTTG	GTCCCCAGGCTTCTAAGTACC
<i>MAP2K1</i> exon 3	CATCCCTCCTCCCTCTTC	AGGCTGAGAGGGTGTACAT
<i>MAP2K1</i> exon 3	TCATCCCTCCTCCCTCTTT	CTCTTAAGGCCATTGCTCCA
<i>U2AF1</i> exon 2	CACAAATGGAAAATACAAC TACGA	TCCCAGCAAAATAATCAGCTC
<i>BRAF</i> exon 15	CAGATATATTCCTCATGAAGACCTCACAGT	TGGGACCCACTCCATCGA
<i>BRAF</i> reporters	CTAGCTACAGTGAAATC	TAGCTACAGAGAAATC

Supplementary Figure 1: Tumor and Normal Sample Purity Estimation.

Alternate read fraction (i.e. fraction of reads supporting non-reference allele) in both tumor (vertical axis) and paired normal (horizontal axis) samples at positions identified as non-reference in either sample. For statistical robustness positions were filtered for those containing a minimum depth of coverage of 50 reads.

